Introduction to Probability and Statistics

Sec 3: Discrete Random Variables

Random Variables

A **Random Variable (RV)** is a variable which assumes numerical values associated with the random outcomes of an experiment. One (and only one) numerical value is assigned to each performance of the experiment.

For example:

- Number of cars crossing the Wouri bridge daily.
- Highest daily temperature in Limbe.
- Number of defective items in a batch.
- Exchange rate of EUR to USD.

Types of random variables

As the case was with types of data, we also have two types of (numerical) random variables:

- Discrete random variables can only assume a finite or countable number of values.
- Continuous random variables can assume values corresponding to any of the points contained in one or more intervals.

In this unit, we concentrate on discrete RVs.

Discrete probability distribution

The probability distribution of a discrete random variable X specifies the probability $p(x)$ for each possible value of X. We write

$$p(x) = \mathbb{P}(X = x).$$

Note: Uppercase X denotes the RV, lowercase x are the values X can take.

- The probability distribution function of a discrete RV is also known as the probability mass function of the RV.

For any discrete probability distribution, the following must be true:

1. $0 \leq p(x) \leq 1$ for all x.
2. $\sum_x p(x) = 1$, where the summation is over all possible values of x.

Examples of Discrete probability distributions

We can use a formula, a table or a graph to represent a probability distribution.

X: Number observed after rolling a fair die.

Table:

x	1	2	3	4	5	6
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6

Graph:

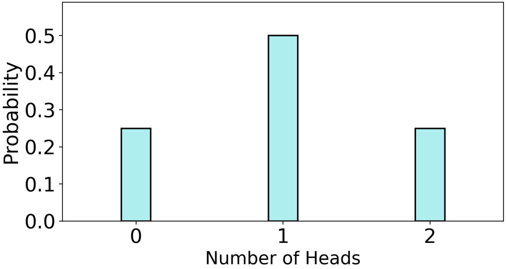
Examples of Discrete probability distributions

Y: Number of heads observed after tossing two coins.

Table:

y	0	1	2
$p(y)$	$1/4$	$1/2$	$1/4$

Graph:



Cumulative distribution: Definition

- The **cumulative distribution function** of discrete random variable X is given by the cumulative probabilities $F(x)$:

$$F(x) = \mathbb{P}(X \leq x) = \sum_{u=0}^{x} p(u)$$

- We use **cdf** in short to refer to the cumulative distribution function.

- The cdf at a point x tells us how much "mass" there is up to x in the distribution of the rv X.

Cumulative distribution: Examples

1. X: Number observed after rolling a fair die:

x	1	2	3	4	5	6
$F(x)$	$1/6$	$2/6$	$3/6$	$4/6$	$5/6$	1

2. Y: Number of heads observed after tossing two coins:

y	0	1	2
$F(y)$	$1/4$	$3/4$	1

Expected Value of Discrete RVs

- The concept of expected value of a discrete RV is strictly linked to the concept of sample mean for "grouped" data: think of Q7 (cars in a car park) or Q8 (the weird-titled sold books) in the tutorial sheet.

- For a sample where value x_i appears with absolute frequency F_i, we have:

$$\bar{x} = \frac{1}{n} \sum_i F_i \, x_i = \sum_i f_i \, x_i \, ,$$

where f_i are the relative frequencies (which sum up to 1).

Expected Value of Discrete RVs

We do exactly the same with the expectation of a RV X:
we take the average of all the values X can take (x_i), weighing each of them by how likely that value is ($p(x_i)$).

- The **mean** or **expected value** of a discrete random variable is:

$$\mu = \mathbb{E}[X] = \sum_i p(x_i)\, x_i = \sum_x p(x)\, x\,.$$

- Notice the parallel with the mean for grouped data

$$\bar{x} = \sum_i f_i\, x_i\,.$$

Expected Value of Discrete RVs: Examples

1. X: Number observed after rolling a fair die:

$$\mathbb{E}[X] = \sum_x p(x)\,x = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \ldots + \frac{1}{6} \times 6 = 3.5$$

2. Y: Number of heads observed after tossing two coins:

$$\mathbb{E}[Y] = \sum_y p(y)\,y = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1$$

Expected Value of Discrete RVs

- Similarly, we can consider the expected value of a *function* of X, g(X):

$$\mathbb{E}[g(X)] = \sum_x p(x)\, g(x)\,.$$

- From the die example, let $g(x) = x^2$. Then:

$$\mathbb{E}\left[X^2\right] = \sum_x p(x)\, x^2 = \frac{1}{6} \times 1^2 + \frac{1}{6} \times 2^2 + \ldots + \frac{1}{6} \times 6^2 = 15.17$$

Properties of Expectation: Intuition

A lot (actually, *almost all*) we see in this chapter has a perfect analogy with what we saw in the first chapter, on samples of data.

For example:

- If x is a costant sample, then its mean is just that constant:

$$x = (3, 3, 3, 3, 3, 3, 3) \quad \to \quad \bar{x} = 3$$

- If you have a sample and multiply it by a constant c, then the mean of the new sample is just c times the mean of the old sample:

$$x = (4, 1, \ 6\ , \ 4, \ 5\ ) \quad \to \quad \bar{x} = 4$$

$$y = (8, 2, 12, 8, 10) \quad \to \quad \bar{y} = 8 \quad (y = 2x)$$

Properties of Expectation: Intuition

- If you have two samples (of equal size) and sum them term by term, then the mean of the sum is the sum of the two means:

$$x = (4, 3, 6, 1, 6) \quad \rightarrow \quad \bar{x} = 4$$

$$y = (5, 2, 1, 4, 3) \quad \rightarrow \quad \bar{y} = 3$$

$$z = (9, 5, 7, 5, 9) \quad \rightarrow \quad \bar{z} = 7 \quad (z = x + y)$$

Properties of Expectation

The following are key properties of the expected value (or expectation) of RVs.

1. If X is a constant RV ($X = c$ for some constant c), then its expected value is c:
$$\mathbb{E}[c] = c.$$

2. If X is a RV and c is a constant, then:
$$\mathbb{E}[cX] = c\,\mathbb{E}[X]$$

3. If X and Y are two RVs, then:
$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Properties of Expectation: Examples

More generally, we say that the expectation is linear:

- It "respects" sums
- Constant can "come out" of the expectation.

Examples

- $$\begin{aligned}\mathbb{E}[3X + 2Y] &= \mathbb{E}[3X] + \mathbb{E}[2Y] \\ &= 3\,\mathbb{E}[X] + 2\,\mathbb{E}[Y]\end{aligned}$$
- $$\mathbb{E}\left[3X^2 - 2X + 5\right] = 3\mathbb{E}\left[X^2\right] - 2\mathbb{E}[X] + 5$$

Variance of a Discrete Random Variable

- **Recall**: The variance of a sample is the "average" squared difference between a value and the sample mean.

- Similarly, the variance of a random variable X is the average squared difference between its values and the mean $\mu = \mathbb{E}[X]$.

- The **variance** of a discrete random variable X is defined as:

$$\sigma^2 = \mathbb{E}\left[(X - \mu)^2\right] = \sum_x p(x)\,(x - \mu)^2$$

- An equivalent and sometimes more convenient formula is:

$$\begin{aligned}\sigma^2 &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= \mathbb{E}\left[X^2\right] - \mu^2\end{aligned}$$

Properties of Variance

The following properties of the variace hold.
Compare them with the expectation properties on slide 14.

1. If X is constant, then its variance is zero:

$$\mathbb{V}\mathrm{ar}\,[c] = 0.$$

2. If X is a random variable and c is a constant, then:

$$\mathbb{V}\mathrm{ar}\,[c\,X] = c^2\,\mathbb{V}\mathrm{ar}\,[X]$$

3. If X and Y are **independent** random variables, then:

$$\mathbb{V}\mathrm{ar}\,[X + Y] = \mathbb{V}\mathrm{ar}\,[X] + \mathbb{V}\mathrm{ar}\,[Y].$$

The square root the variance of a rv X is the **standard deviation** of X.

Properties of Mean and Variance

Overall, if we have any two random variables X and Y, and two constants a and b, we can say that:

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

$$\mathbb{V}\text{ar}\,[aX + b] = a^2\,\mathbb{V}\text{ar}\,[X]$$

Equivalently, by taking the square root of the last equation:

$$\text{sd}\,[aX + b] = |a|\;\text{sd}[X]\,.$$

Example: Rolling one Die

- Let X be the number displayed after rolling one die.
- From previous calculations we know that:
  - $\mathbb{E}[X] = 3.5$
  - $\mathbb{E}\left[X^2\right] = 15.17$
- What is the standard deviation of X?

$$\begin{aligned}\mathbb{V}\text{ar}\,[X] &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\ &= 15.17 - 3.5^2 = 2.92\end{aligned}$$

$$\Longrightarrow \quad \text{sd}(X) = \sqrt{2.92} = 1.71.$$

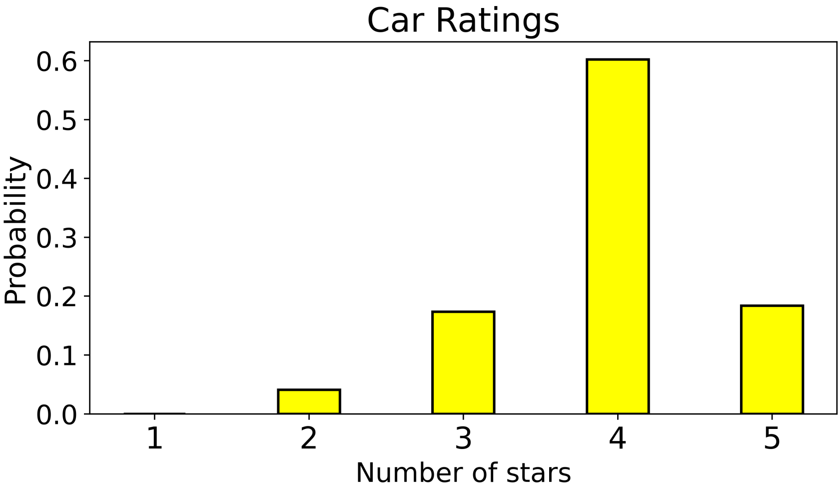
Example: Car Ratings

98 cars were tested and rated on a scale of 1-5 stars with regard to their safety level in a collision. The star allocations are summarised in the following table:

★	★★	★★★	★★★★	★★★★★
0	4	17	59	18

- Calculate the probability distribution for Y: the number of stars obtained by selecting one of the cars at random.
- Calculate $\mathbb{P}(Y \leq 3)$.
- Hence calculate $\mathbb{P}(Y > 3)$.
- What is the expected value of Y?
- Calculate the standard deviation of Y.

Example: Car Ratings



Common Discrete Distributions

Common Examples of Discrete Distributions

After reviewing general properties of discrete RVs, we now study three common discrete probability distributions and the context where they arise.

These are:

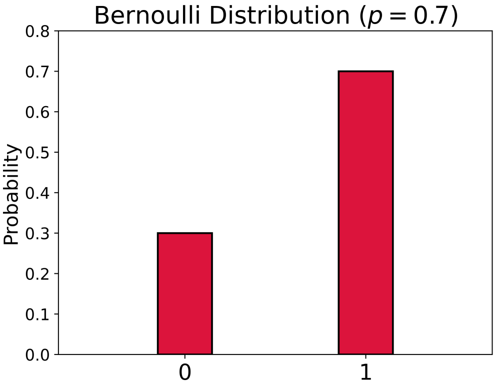
1. Bernoulli Distribution
2. Binomial Distribution
3. Poisson Distribution

Bernoulli Distribution

- Let's consider a very simple RV X, which takes the value 1 with probability p, and 0 with probability $1 - p$.

Probability mass function

x	0	1
$p(x)$	$1 - p$	p



Bernoulli Distribution

❖ What are the mean and variance of such random variable?

$$\begin{aligned}
\mathbb{E}[X] &= (1-p)\cdot 0 + p\cdot 1 = p \\
\mathbb{V}\mathrm{ar}[X] &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X] - \mathbb{E}[X]^2 \qquad (\text{since } X = X^2!) \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}$$

Such a rv is called a **Bernoulli random variable with parameter p**. We write:

$$X \sim \mathcal{B}er(p)$$

and we have:

$$X \sim \mathcal{B}er(p) \quad \Rightarrow \quad \begin{cases} \mathbb{E}[X] = p \\ \mathbb{V}\mathrm{ar}[X] = p(1-p). \end{cases}$$

Binomial Distribution

Binomial Distribution: Example

Suppose we toss a coin 8 times. At each toss, the probability of getting head (H) is 0.7, the one of getting tail (T) is 0.3. What's the probability of getting 5H and 3T?

Let's break this down.

- What's the probability of getting the following particular sequence?

  H T H H T T H H

  That's:

  $$0.7 \cdot 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.3 \cdot 0.3 \cdot 0.7 \cdot 0.7$$

- And the probability of this other sequence?

  T T H H H H T H

  $$0.3 \cdot 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.7 \cdot 0.7 \cdot 0.3 \cdot 0.7$$

Binomial Distribution: Example

- In fact, **any** specific sequence with 5 Hs and 3 Ts will have probability:

$$0.7^5 \times 0.3^3$$

- The question is then: How many such sequences are there?
- We need to count in how many ways we can select the 5 heads positions out of the 8:

?		?		?		?	?

- Bu we learnt this is $\binom{8}{5} = 56$ (remember the fruit?!)
- So, the probability of getting 5 heads out of 8 tosses is:

$$\mathbb{P}(\{5 \text{ heads out of } 8\}) = \binom{8}{5} \cdot 0.7^5 \cdot 0.3^3 = 0.254.$$

Binomial Distribution

More generally, we have shown the following. If:

- An experiment is conducted N times ($N = 8$), with each trial independent of the other.

- Each trial has probability p of being a "success" ($p = 0.7$), and probability $q = 1 - p$ of being a "failure" ($q = 0.3$).

Then:

- The probability of obtaining exactly k successes out of N is:

$$\mathbb{P}(\{\text{k successes}\}) = \binom{N}{k} \cdot p^k \cdot q^{N-k}.$$

Binomial Distribution

Let X be the random variable which counts the number of successes in $\boldsymbol{N}$ independent trials, where each trial has probability $\boldsymbol{p}$ of resulting in a success.

Then, X follows a Binomial distribution with parameters N and p. We write:

$$X \sim \mathcal{B}(N, p).$$

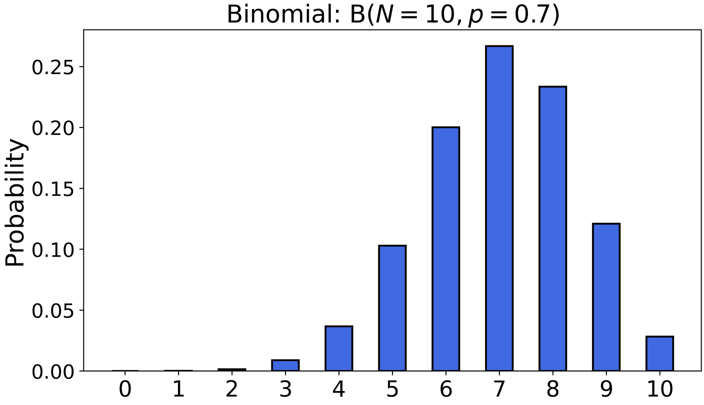
The probability mass function of X is:

$$\mathbb{P}(X = k) = \binom{N}{k} \cdot p^k \cdot q^{N-k},$$

where $q = 1 - p$.

Binomial Distribution: Interactive Visualisation

Python Notebook or Geogebra Link 1 or Geogebra Link 2.



Binomial Distribution: Example 2

A student takes 20 short multiple-choice quizzes.

- On each quiz, the probability of answering correctly is 0.9.

What is the probability that the student answers at least 18 correctly?

Let X be the number of correct answers. $X \sim \mathcal{B}(20, 0.9)$. Hence:

$$\begin{aligned}
\mathbb{P}(X \geq 18) &= \mathbb{P}(X = 18) + \mathbb{P}(X = 19) + \mathbb{P}(X = 20) \\
&= \binom{20}{18} 0.9^{18}\, 0.1^{2} + \binom{20}{19} 0.9^{19}\, 0.1^{1} + \binom{20}{20} 0.9^{20}\, 0.1^{0} \\
&= 190 \cdot 0.9^{18}\, 0.1^{2} + 20 \cdot 0.9^{19}\, 0.1 + 0.9^{20} \\
&= 0.677
\end{aligned}$$

Binomial Distribution: Mean and Variance

What are the mean and variance of a Binomial random variable $X \sim \mathcal{B}(N, p)$?

Idea

- For each of the N experiments, let's consider the rv X_i which records a 1 if a success happened in experiment i, and a 0 if a failure happened in experiment i.
- Then, the number of successes X is simply

$$X = X_1 + X_2 + \cdots + X_N.$$

We have that:

- $X_i \sim \mathcal{B}er(p)$
- All X_i's are independent of each other.

Binomial Distribution: Mean and Variance

Recalling the mean and variance of a Bernoulli RV (p and $p(1-p)$, respectively), and the properties of expectation and variance, we get:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_N] \\ &= p + \cdots + p \\ &= Np\end{aligned}$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}\,[X] &= \mathbb{V}\mathrm{ar}\,[X_1] + \cdots + \mathbb{V}\mathrm{ar}\,[X_N] \quad \text{(due to independence!)} \\ &= p(1-p) + \cdots + p(1=p) \\ &= Np(1-p)\end{aligned}$$

So:

$$X \sim \mathcal{B}(N,p) \quad \Rightarrow \quad \begin{cases} \mathbb{E}[X] = Np \\ \mathbb{V}\mathrm{ar}\,[X] = Np(1-p). \end{cases}$$

Recap: Bernoulli and Binomial

Bernoulli Distribution

- Only two values: 0 and 1 (with probability $1-p$ and p, respectively)
- $$\mathbb{E}[X] = p, \quad \mathbb{V}\mathrm{ar}[X] = p(1-p).$$

Binomial Distribution

- Counts the number of successes in N independent trials, where each trial results in a success with probability p.
- $$\mathbb{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$
- $$\mathbb{E}[X] = Np, \quad \mathbb{V}\mathrm{ar}[X] = Np(1-p).$$

Poisson Distribution

Poisson Random Variable: Examples

Consider the following examples of random variables.

- The number of vehicles crossing a bridge in a day.
- The number of jobs sent to a computer processor in an hour.
- The number of road signs in 500 metres of road.
- The number of pollutant particles in a litre of river water.

All of these involve **counting events** in a fixed amount of time, space, or area. Typically, a **Poisson** random variable is used for these situations.

Poisson Distribution: Assumptions

To model counts with a Poisson distribution, three conditions should hold:

- Events occur **independently**. (One event does not influence another)
- In a very short interval, **at most one** event can occur.
  (Two events cannot occurr exactly at the same time)
- Events occur at a **constant average rate** λ.
  (On average, λ events per unit of time)

Under these assumptions, the number X of events in the interval follows a **Poisson distribution with rate λ**. We write:

$$X \sim \mathcal{P}oi(\lambda),$$

where λ is the number of events that, on average, we expect in that interval.

Poisson Distribution: Formulas

If $X \sim \mathcal{P}oi(\lambda)$, then the probability of observing exactly n events is:

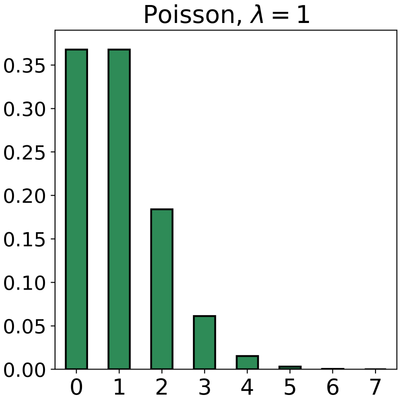
$$\boxed{\mathbb{P}(X = n) = \frac{e^{-\lambda}\,\lambda^n}{n!}, \qquad n \in \mathbb{N}}$$

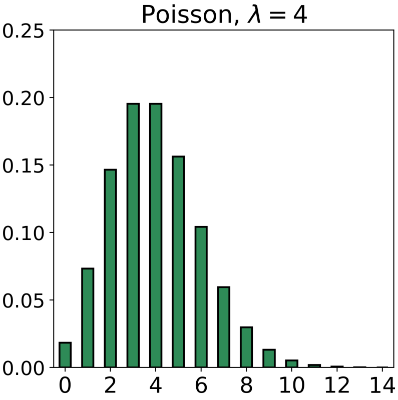
As we said, λ represents the average number of events that happen in the interval (the mean of X). However, it also represents the variance of X:

$$X \sim \mathcal{P}oi(\lambda) \quad \Rightarrow \quad \begin{cases} \mathbb{E}[X] = \lambda \\ \mathbb{V}\mathrm{ar}[X] = \lambda. \end{cases}$$

Poisson Distribution: Interactive Visualisation

Python Notebook or Geogebra Link.





Poisson Distribution: Example

On average, 4 taxis pass in front of AIMS every ten minutes. Between 11:00 and 11:10:

1. What is the probability that exactly five taxis will pass?
2. What is the probability that at most two taxis will pass?

Solution. Let X be the number of taxis in the 10-minute interval. Then:

$$X \sim \mathcal{P}oi(\lambda = 4).$$

1. Exactly five taxis:

$$\mathbb{P}(X = 5) = \frac{e^{-\lambda}\,\lambda^n}{n!} = \frac{e^{-4}\,4^5}{5!} \approx 0.156.$$

2. At most two taxis:

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = e^{-4} \cdot 13 \approx 0.238.$$

More on Poisson Distribution

What if, under the same assumptions as before (average of 4 taxis in ten minutes), we ask:

- What is the probability that at least ten taxis will pass *between 11:00 and 11:30*?

Let Y be the number of taxis passing between 11:00 and 11:30. Then:

$$Y \sim \mathcal{P}oi(12)\,.$$

Therefore:

$$\begin{aligned} \mathbb{P}(Y \geq 10) &= 1 - \mathbb{P}(Y \leq 9) \\ &= 1 - \sum_{n=0}^{9} e^{-12} \frac{12^n}{n!} \approx 0.7576\,. \end{aligned}$$

Poisson Distribution: Property

- Consider two Poisson RVs, X and Y, with rates λ and μ.
- If X and Y are independent, then $X + Y$ is also Poisson, with rate $\lambda + \mu$.

$$\begin{cases} X \sim \mathcal{P}oi(\lambda) \\ Y \sim \mathcal{P}oi(\mu) \end{cases} \text{independent} \quad \Longrightarrow \quad X + Y \sim \mathcal{P}oi(\lambda + \mu)$$

Poisson Process

Suppose events occur following a Poisson law at the rate of λ per unit of time. Then the number of events occurred up to time t, X_t, still follows a Poisson distribution, but with rate λt:

$$X_t \sim \mathcal{P}oi(\lambda t)$$

This is called a Poisson Process with rate λ.